

Appendix

Appendix A1.1 Study characteristics: The San Ramon Study (randomized controlled trial with confounding problems)¹

Characteristic	Description
Study citation	Battistich, V., Solomon, D., Watson, M., Solomon, J., & Schaps, E. (1989). Effects of an elementary program to enhance prosocial behavior on children's cognitive-social problem-solving skills and strategies. <i>Journal of Applied Developmental Psychology</i> , 10(2), 147–169.
Participants	The study included two cohorts of elementary school students from six elementary schools. Each cohort began with the project in kindergarten. The first cohort started kindergarten in 1982–83 and the second in 1985–86. The first cohort included 13 kindergarten classrooms that were followed through the elementary school years; ² the second cohort included 14 classrooms that were followed through the first grade. There were about 350 students a year in the first cohort (divided evenly between the intervention and comparison groups). Of those 350 students, about 165 students remained with the cohort all five years.
Setting	The study took place in six elementary schools located in a middle- to upper middle-class suburban community in northern California.
Intervention	The intervention schools implemented the <i>Child Development Project</i> (CDP) program. (For details about the connection between the CDP and the CSC, see the CSC intervention report). ³ Students in the intervention group received the CDP program every year starting in kindergarten. Class meetings in the intervention condition included activities designed to promote core values. In the classroom, students learned group interaction skills and relevant values and worked in small groups toward mutual academic and nonacademic goals. Teachers identified and discussed exemplary behavior using examples from the classroom, television, literature, and movies. Developmental discipline, a classroom management approach, was applied to teach prosocial norms and values. In addition, children were encouraged to help others by doing classroom chores, tutoring younger students as part of the “buddies” programs, performing charitable community activities, and helping with activities in the school at large. An implementation check done by two independent observers indicated a high level of implementation and significantly different classroom experiences (with respect to classroom activities and teacher behavior) in the intervention classrooms compared with the comparison classrooms.
Comparison	The comparison group included three elementary schools in the same school district as the intervention schools and matched with the intervention schools on socioeconomic status and interest in the intervention. Comparison group students did not participate in the <i>Caring School Community</i> program. No information was provided on character education related practices in the comparison schools.
Primary outcomes and measurement	Student outcomes in three domains were examined: behavior; knowledge, attitudes, and values; and academic achievement. Students' behavior was assessed using direct observations of students' behavior in the classroom. Students' knowledge, attitudes, and values were assessed using several self-report questionnaires. Academic achievement was assessed using standardized achievement tests. (See Appendices A2.1–A2.3 for more detailed descriptions of the outcome measures.)
Teacher training	Teacher training consisted of a one-week summer institute, monthly workshops, frequent meetings with project staff who also observed the classrooms periodically, and supporting curriculum materials.

1. The San Ramon Study randomly assigned one group of three schools to intervention or comparison groups. Because the unit of assignment consisted of one set of schools, there is confounding between the unit of assignment and the unit of intervention. The study authors collected baseline measures and demonstrated that the intervention and comparison schools were well matched in terms of relevant students' outcomes. Therefore, although this study did not meet WWC standards as a randomized controlled trial because of the confounding effect, it met standards as a quasi-experimental design.
2. The study authors also conducted follow-up analyses to examine student outcomes in grades 5 and 6 (Battistich, 2003; Solomon, Battistich, & Watson, 1993; Solomon, Watson, Battistich, Schaps, & Delucchi, 1996); however, in grade 5 only one school per condition participated in the study, therefore, due to confounding of the school with the intervention, this analysis was not reviewed. In addition, in grade 6, only two schools per condition participated in the analysis; therefore, due to severe attrition on the school level, and small percentage of students receiving the program since kindergarten (35%) this follow-up analysis was not reviewed.
3. According to Battistich (2003), this intervention has recently been modified in a significant way. Therefore, this study presents an assessment of the impact of the intervention as it was configured at the time of the study.

Appendix A1.2 Study characteristics: The Six-District Study (quasi-experimental design)

Characteristic	Description
Study citation	Battistich, V., Schaps, E., Watson, M., Solomon, D., & Lewis, C. (2000). Effects of the Child Development Project on students' drug use and other problem behaviors. <i>Journal of Primary Prevention</i> , 21(1), 75–99.
Participants	Participants of the study were students in the upper elementary grades in 12 intervention schools and 12 matched comparison schools in six districts (grades 3–5 in four districts and grades 4–6 in the two other districts). This review includes only five intervention schools with meaningful progress toward program implementation and their matched comparison schools. ¹ The composition of the student population was similar at the intervention and comparison schools. Two of the schools in the sample reviewed and their matched comparison schools served a predominantly low-socioeconomic status population. In four pairs of schools, most of the school population was white; in one pair of schools, most of the students were African-American. The students began with the study in 1991–92 when they were in the third or fourth grade and were followed until the end of elementary school. ²
Setting	The study took place in 24 schools located in six urban, suburban, and rural districts and serving diverse student populations. The 10 schools included in this review were from three districts: one district on the West Coast, one in the South, and one in the Southeast. Two schools (one intervention and one comparison) were from a rural school district. Four schools (two intervention and two comparison) were located in an urban school district, and four schools (two intervention and two comparison) were located in a suburban school district.
Intervention	The intervention schools implemented the <i>Child Development Project</i> (CDP) program. (For details about the connection between the CDP and the CSC, see the CSC intervention report). The CDP program consisted of classroom discussions and activities, a schoolwide component, and a family involvement component. Class meetings included activities designed to promote core values. In the classrooms, students learned group interaction skills and relevant values and worked in small groups toward mutual academic and nonacademic goals. Teachers identified and discussed exemplary behavior using examples from the classroom, television, literature, and movies. Developmental discipline, a classroom management approach, was applied to teach prosocial norms and values. In addition, children were encouraged to help others by doing classroom chores, tutoring younger students as part of the “buddies” programs, performing charitable community activities, and helping with activities in the school at large. Classroom observations and interviews with school staff indicated an adequate level of program implementation.
Comparison	The comparison schools were drawn from the same school districts as the intervention schools and matched with the intervention schools with respect to school size and student characteristics. The comparison schools did not implement the program.
Primary outcomes and measurement	The study investigated students' drug use and other types of problem behavior, core values (acceptance of people in outgroups, concern for others, altruistic behavior), and academic attitudes and motives (sense of the school as a community, task orientation, frequency of reading self-chosen books outside of school, frequency of reading self-chosen books in school, enjoyment of class, preference for challenging tasks). (See Appendices A2.1–A2.3 for more detailed descriptions of the outcome measures.)
Teacher training	Professional development was conducted at both the district and the school levels. At first, the program was introduced to 8–15 member “implementation teams” in each district. In the three subsequent years of the study, schoolwide training was also conducted. Each year, the implementation teams participated in summer workshops delivered by the developer. Implementation team members took increasing responsibility for the within-district workshops and for other support to teachers implementing the program. Teachers were also encouraged to meet regularly in small “partner study and support groups” to discuss and help each other with implementation issues.

1. The other seven intervention schools in four school districts did not demonstrate “meaningful progress towards implementation of the CDP program,” according to the study authors, with as many teachers in these schools showing no changes or even declines from baseline on a measure of program implementation as the number of teachers showing positive changes from baseline in implementation. Therefore, these seven intervention schools and their matched comparison schools were not included in this review.
2. The study authors (Battistich, 2001; Battistich, Schaps, & Wilson, 2004) conducted additional follow-up analyses when the students attended middle schools. But these analyses did not identify those students from feeder schools that reached meaningful progress toward implementation. In addition, these analyses did not control for baseline differences between students in the intervention and comparison groups; therefore, they were not included in this review.

Appendix A2.1 Outcome measures in the behavior domain

Outcome measure	Description
Negative behavior	Students' interpersonal negative behavior in the classroom. This is scale derived from a classroom observation sign-system instrument developed for the purposes of the study reported by Battistich et al. (1989). Segments of two-minute observations were coded using a prespecified rating system to indicate frequency and quality of various behaviors and activities.
Spontaneous prosocial behavior	Students' spontaneous helpfulness, concern for others, and cooperation. This is a scale derived from a classroom observation sign-system instrument developed for the purposes of the study reported by Battistich et al. (1989). Segments of two-minute observations were coded using a prespecified rating system to indicate frequency and quality of various behaviors and activities.
Harmoniousness	Students' harmoniousness, apparent interest and involvement, and apparent happiness. This is a scale derived from a classroom observation sign-system instrument developed for the purposes of the study reported by Battistich et al. (1989). Segments of two-minute observations were coded using a prespecified rating system to indicate frequency and quality of various behaviors and activities.
Supportive, friendly, and helpful behavior	Students' supportive, friendly, and helpful behavior in the classroom, which may include one or more of the following behaviors: support and encouragement to other students, affection, inviting others to join activities, and thanking or praising other students. This is a scale derived from a classroom observation sign-system instrument developed for the purposes of the study reported by Battistich et al. (1989). Segments of two-minute observations were coded using a prespecified rating system to indicate frequency and quality of various behaviors and activities.
Social competence	Teacher's ratings of students' social competence based on students' behavior (as cited in Solomon et. al., 1989). Four ratings were combined to form this general measure (takes an active role in resolving personal difficulties or problems; is admired and sought after by peers; does not expect others to provide for his or her every need; and gets along easily and comfortably with adults). Teachers rated each student on his or her social competence relative to other students in the class.
Altruistic behavior	A 10-item student self-report measure derived from Rushton, Chrisjohn, and Fekken (1981; as cited in Solomon et al., 2000) that assesses students' altruistic behavior.
Use of alcohol	A single survey item assessing lifetime use of alcohol (as cited in Battistich et al., 2000).
Use of marijuana	A single survey item assessing lifetime use of marijuana (as cited in Battistich et al., 2000).
Use of cigarettes	A single survey item assessing lifetime use of cigarettes (as cited in Battistich et al., 2000).
Ran away from home	A single survey item assessing frequency of involvement in one type of delinquent behavior: ran away from home during the past year (as cited in Battistich et al., 2000).
Skipped school	A single survey item assessing frequency of involvement in one type of delinquent behavior: skipped school during the past year (as cited in Battistich et al., 2000).
Damaged property on purpose	A single survey item assessing frequency of involvement in one type of delinquent behavior: damaged property on purpose during the past year (as cited in Battistich et al., 2000).
Stolen money or property	A single survey item assessing frequency of involvement in one type of delinquent behavior: stolen (or attempted to steal) money or property during the past year (as cited in Battistich et al., 2000).
Carried a knife, gun, or other weapon	A single survey item assessing frequency of involvement in one type of delinquent behavior: carried a knife, gun, or other weapon during the past year (as cited in Battistich et al., 2000).
Threatened to hurt someone	A single survey item assessing frequency of involvement in one type of delinquent behavior: threatened to hurt someone during the past year (as cited in Battistich et al., 2000).
Hurt someone on purpose	A single survey item assessing frequency of involvement in one type of delinquent behavior: hurt someone on purpose during the past year (as cited in Battistich et al., 2000).

(continued)

Appendix A2.1 Outcome measures in the behavior domain *(continued)*

Outcome measure	Description
Taken a car without permission	A single survey item assessing frequency of involvement in one type of delinquent behavior: taken a car without permission during the past year (as cited in Battistich et al., 2000).
Been in a gang fight	A single survey item assessing frequency of involvement in one type of delinquent behavior: been in a gang fight during the past year (as cited in Battistich et al., 2000).
Thrown objects at people	A single survey item assessing frequency of involvement in one type of delinquent behavior: thrown objects at people during the past year (as cited in Battistich et al., 2000).
Been made fun of or called names¹	A single survey item assessing frequency of being the subject of one type of victimization behavior: been made fun of or called names during the past year (as cited in Battistich et al., 2000).
Had property damaged on purpose¹	A single survey item assessing frequency of being the subject of one type of victimization behavior: had property damaged on purpose during the past year (as cited in Battistich et al., 2000).
Had property stolen from desk¹	A single survey item assessing frequency of being the subject of one type of victimization behavior: had property stolen from desk during the past year (as cited in Battistich et al., 2000).
Had money or property taken by force or threat¹	A single survey item assessing frequency of being the subject of one type of victimization behavior: had money or property taken by force or threat during the past year (as cited in Battistich et al., 2000).
Been threatened with harm¹	A single survey item assessing frequency of being the subject of one type of victimization behavior: been threatened with harm during the past year (as cited in Battistich et al., 2000).
Been physically attacked¹	A single survey item assessing frequency of being the subject of one type of victimization behavior: been physically attacked during the past year (as cited in Battistich et al., 2000).

1. This victimization measure is an indicator of students' problem behavior inside school.

Appendix A2.2 Outcome measures in the knowledge, attitudes, and values domain¹

Outcome measure	Description
Democratic values	A total score of several democratic values subscales. The subscales measure students' endorsement of statements favoring equality of representation and participation, willingness to compromise, and belief in one's responsibility to state opinions, even if unpopular.
Perceptual benevolence	A self-report measure of the prosocial value of kindness to others (as cited in Benninga et al., 1992).
Empathy	An 11-item self-report paper-and-pencil measure of empathy that was adapted from Bryant (1982; as cited in Solomon et al., 1996).
Concern for others	This measure asks students to indicate their views regarding others' problems and whether they feel empathy toward others and their problems. This measure was adapted from Solomon and Kendall (1979; as cited in Solomon et al., 1996).
Concern for equality	A self-report measure that addresses students' prosocial concern for equality of participation and outcomes for individuals in social situations (as cited in Benninga et al., 1992).
Motive to help others learn	A self-report measure that assesses students' motivation in helping other students in academic learning (as cited in Benninga et al., 1991).

(continued)

Appendix A2.2 Outcome measures in the knowledge, attitudes, and values domain *(continued)*

Outcome measure	Description
Social understanding	This six-item interview was adapted from Flapan (1968; as cited in Solomon et al., 1996). Students responded to a series of questions about scenes from a movie (<i>Our Vines Have Tender Grapes</i> ; MGM, 1945) showing a series of interactions between two children and two adults.
Conflict resolution interview: general conflict resolution strategy	A score derived from an individually administered interview that presents students with three hypothetical conflict situations and prompts proposed conflict resolution strategies (as cited in Battistich et. al., 1989). This outcome addresses students' suggested strategies to resolve social problems (for example, using aggression, cooperation, or appealing to authority), taking into account who is favored by this strategy (self, other, or both) and the needs of self, other, or both.
Social problem solving: interpersonal sensitivity	A score derived from the Social Problem-Solving Analysis Measure (SPSAM) (Elias et al., 1978; as cited in Battistich et. al., 1989). This outcome addresses students' understanding of social problem situations and awareness of the thoughts and feelings of the person involved in each situation.
Social problem solving: means-ends cognitive problem solving	A score derived from the SPSAM by Elias et al. (1978; as cited in Battistich et. al., 1989). This outcome addresses students' ability to plan specific steps toward resolution, consider alternative courses of action, and anticipate obstacles to and consequences of one's action.
Social problem solving: obstacle means-end cognitive problem solving	A score derived from the SPSAM by Elias et al. (1978; as cited in Battistich et. al., 1989). This outcome addresses students' ability to plan specific steps toward resolution, consider alternative courses of action, and anticipate consequences and additional obstacles when challenged or confronted by an obstacle.
Social problem solving: obstacle outcome expectancies	A score derived from the SPSAM by Elias et al. (1978; as cited in Battistich et. al., 1989). This outcome addresses students' belief that a strategy they proposed as an amendment to previously challenged suggestions will lead to a successful resolution of the interpersonal problem.
Social problem solving: obstacle problem resolution strategies	A score derived from the SPSAM by Elias et al. (1978; as cited in Battistich et. al., 1989). This outcome addresses students' suggested strategy to resolve social problems (for example, using aggression, cooperation, or appealing to authority) when their previously proposed strategies are challenged.
Social problem solving: outcome expectancies	A score derived from the SPSAM by Elias et al. (1978; as cited in Battistich et. al., 1989). This outcome addresses students' belief that certain actions will lead to a successful resolution of the interpersonal problem.
Social problem solving: primary resolution strategies	A score derived from the SPSAM by Elias et al. (1978; as cited in Battistich et. al., 1989). This outcome addresses students' suggested strategy to resolve social problems (for example, using aggression, cooperation, or appealing to authority), take into account who is favored by this strategy (self, other, or both), and consider the needs of self, other, or both.
Social problem solving: proportion antisocial strategies	A score derived from the SPSAM, an individual interview by Elias et al. (1978; as cited in Battistich et. al., 1989). This outcome addresses students' consideration and relative use of antisocial strategies such as an aggressive or disruptive action.
Social problem solving: proportion prosocial strategies	A score derived from the SPSAM by Elias et al. (1978; as cited in Battistich et. al., 1989). This outcome addresses students' consideration and relative use of prosocial strategies such as cooperation and sharing, polite request, and discussion of the problem.
Conflict resolution interview: proportion antisocial strategies	A score derived from the conflict resolution interview (as cited in Battistich et. al., 1989). This individual interview presents students with three hypothetical conflict situations and prompts proposed conflict resolution strategies. This outcome addresses students' consideration and relative use of such antisocial strategies as an aggressive or disruptive action.
Conflict resolution interview: proportion prosocial strategies	A score derived from the conflict resolution interview (as cited in Battistich et al., 1989). This individual interview presents students with three hypothetical conflict situations and prompts proposed conflict resolution strategies. This outcome addresses students' consideration and relative use of prosocial strategies such as cooperation and sharing, polite request, and discussion of the problem.

(continued)

Appendix A2.2 Outcome measures in the knowledge, attitudes, and values domain *(continued)*

Outcome measure	Description
Conflict resolution interview: consideration of others' needs	A score derived from the conflict resolution interview (as cited in Battistich et al., 1989). This individual interview presents students with three hypothetical conflict situations and prompts proposed conflict resolution strategies. This outcome addresses students' consideration of the others' needs as well as one's own.
Total self-esteem	A self-report measure (as cited in Benninga et al., 1992; Solomon et al., 1996). For the third-grade level, this score includes general self-esteem (personal feelings of self-worth) and academic self-esteem. For the fourth-grade level, this score also includes social self-esteem.
Sense of community	A self-report measure of students' sense of the classroom and the school as a community that was developed for the purposes of the study reported by Battistich et al. (1989), Solomon et al. (1992), and Solomon et al. (1996). Sense of community is conceptualized as encompassing two main student perceptions: that their behavior and their classmates' behavior show that they care about and are supportive of one another and that they have an important role in classroom decisionmaking and direction. In the study reported by Battistich et al. (2000) sense of community is measured using a 38-item survey that is composed of three subscales: student autonomy and influence in the classroom, classroom supportiveness, and school supportiveness. ²
Acceptance of outgroups	A 10-item student self-report measure that assesses students' acceptance of outgroups (as cited in Solomon et al., 2000). Students indicated how much they would want to do a specific task with various other people (not including a close friend) differing in social distance.
Outgroups discrepancy score (deviation from friend)	A 10-item student self-report measure that assesses students' acceptance of outgroups (as cited in Solomon et al., 2000). Students indicated how much they would want to do a specific task with a close friend and with various other people differing in social distance. The discrepancy score subtracts scores for the members of other groups from that for the friend.
Enjoyment of helping others learn	A five-item student self-report measure adapted from Deer et al. (1988) and Solomon, Watson, Battistich, Schaps, and Delucchi (1992; as cited in Solomon et al., 2000), which assesses students' enjoyment of helping other students learn.
Social competence	A 10-item student self-report measure that assesses social competence (as cited in Solomon et al., 2000).
General self-esteem	A four-item student self-report measure adapted from Solomon and Kendall (1979; as cited in Solomon et al., 2000) that measures student general self-esteem.
Conflict resolution skills	An eight-item student self-report measure adapted from Battistich et al. (1989; as cited in Solomon et al., 2000).
Sense of efficacy	A 10-item student self report measure developed by Cowen, Work, Hightower, Wyman, Pancer, & Lotyczewski (1991; as cited by Solomon et al., 2000).

1. This appendix presents a brief description of all outcomes that are presented in Appendices A3.1–A4.3.

2. Students' sense of autonomy in the classrooms may include elements of teacher behavior in addition to student behavior. The WWC did not obtain statistical information on the classroom supportiveness and school supportiveness subscales separately; therefore the composite score was reviewed.

Appendix A2.3 Outcome measures in the academic achievement domain¹

Outcome measure	Description
California Achievement Test (CAT)	A standardized test that measures achievement in reading, language, spelling, mathematics, study skills, science, and social studies (as cited in Solomon et al., 1992).
Holistic measure of reading comprehension	A measure developed by the Educational Testing Services that assesses the use of high-order thinking and the development of text understanding (as cited in Solomon et al., 1992). Students are asked to read two brief passages and then respond to general questions about the meaning of the passages.
Inductive reasoning	A cognitive ability test adapted from Ennis and Millman (1985; as cited in Battistich et al., 2000) that presents a series of questions to students about the implications of various clues for solving mysteries.
Stanford Achievement Test (SAT9)	A national standardized test that measures student achievement in reading, language, spelling, study skills, listening, mathematics, science, and social science.
SRA Achievement Series	Achievement series that includes two forms, Forms 1 and 2. This series was published by the Science Research Associates, Inc. This test is designed to assess broad areas of knowledge, general skills, and their application.
Liking for reading	One survey item that assesses students' enjoyment of reading books (as cited in Solomon et al., 2000).
Educational aspirations	One survey item that asks students how far they would like to go in school. Students rate their predictions on a five-point scale that ranges from "go to high school, but not graduate" to "finish college."
Educational expectations	One survey item that asks students how far they think they really will go in school. Students rate their predictions on a five-point scale that ranges from "go to high school, but not graduate" to "finish college."
Achievement motivation	A seven-item self-report measure that was used for the study reported by Battistich et al. (1989). The measure was adapted from Weiner and Kukla (1970) and Solomon and Kendall (1979) and assesses achievement motivation at school (as cited in Solomon et al., 1996).
Intrinsic academic motivation	A measure based on a student survey that was used for the study reported by Battistich et al. (2000). This measure is a ratio of two scales: intrinsic and extrinsic academic scales. This scale asks students to indicate why and when they typically do academic work. The measure was adapted from Connell and Ryan (1987) and Deer, Solomon, Watson, and Solomon (1988; as cited in Battistich et al., 2000).
Task orientation	An eight-item self-report measure that assesses the tendency to feel most satisfied when school work is challenging and leads to improved understanding. This measure, used in the study reported by Battistich et al. (2000), was developed by Nicholls (1989; as cited in Solomon et al., 2000).
Ego orientation	A four-item self-report measure that assesses tendency to feel most satisfied when school work allows one to demonstrate better performance than other students. This measure, used in the study reported by Battistich et al. (2000), was developed by Nicholls (1989; as cited in Solomon et al., 2000).
Work avoidance	A five-item measure of a student's tendency to feel most satisfied when work is easy. This measure, used in the study reported by Battistich et al. (2000), was developed by Nicholls (1989; as cited in Solomon et al., 2000).
Preference for challenging tasks	A five-item self-report measure that was adapted from Weiner and Kukla (1970) and Solomon and Kendall (1979; as cited in Solomon et al., 2000).
Frequency reading self-chosen books outside of school	One item rated on a five-point scale that was developed for the purposes of the study reported by Battistich et al. (1989). This item assesses the frequency and enjoyment of reading books outside school.
Frequency reading self-chosen books in school	One item rated on a five-point scale that was developed for the purposes of the study reported by Battistich et al. (1989). This item assesses the frequency and enjoyment of reading books inside school.

1. This appendix presents a brief description of all outcomes that are presented in Appendices A3.1–A4.3.

Appendix A3.1 Summary of study findings included in the rating for the behavior domain¹

			Author's findings from the study					
			Mean outcome (standard deviation ²)		WWC calculations			
Outcome measure ³	Study sample	Sample size (students/ schools)	Caring School Community group	Comparison group	Mean difference ⁴ (Caring School Community – comparison)	Effect size ⁵	Statistical significance ⁶ (at $\alpha = 0.05$)	Improvement index ⁷
The San Ramon Study (randomized controlled trial with confounding problems)								
Negative behavior	Grades K–4	350/6	51.33 (10)	48.36 (10)	2.97	0.29	ns	+12
Spontaneous prosocial behavior	Grades K–4	350/6	53.36 (10)	45.85 (10)	7.51	0.74	Statistically significant	+27
Harmoniousness	Grades K–4	350/6	50.43 (10)	49.47 (10)	0.96	0.09	ns	+4
Supportive, friendly, and helpful behavior	Grades K–4	350/6	52.35 (10)	47.11 (10)	5.24	0.52	Statistically significant	+20
Social competence	Grade 4	295/6	1.66 (0.46)	1.62 (0.47)	0.04	0.09	ns	+3
Average ⁸ for behavior (The San Ramon Study)						0.35	ns	+14
The Six-District Study (quasi-experimental design)								
Altruistic behavior	Grades 3–6	1,986/10	0.04 (0.63)	–0.03 (0.66)	0.07	0.11	ns	+4
Use of alcohol	Grades 5–6	635/10	0.27 (0.39)	0.29 (0.40)	0.02	0.05	ns	+2
Use of marijuana	Grades 5–6	635/10	0.04 (0.22)	0.07 (0.28)	0.03	0.12	ns	+5
Use of cigarettes	Grades 5–6	635/10	0.13 (0.30)	0.12 (0.32)	–0.01	–0.03	ns	–1
Ran away from home	Grades 5–6	635/10	0.09 (0.30)	0.10 (0.34)	0.01	0.03	ns	+1
Skipped school	Grades 5–6	635/10	0.16 (0.38)	0.15 (0.37)	–0.01	–0.03	ns	–1
Damaged property on purpose	Grades 5–6	635/10	0.23 (0.45)	0.27 (0.45)	0.04	0.09	ns	+4

(continued)

Appendix A3.1 Summary of study findings included in the rating for the behavior domain¹ (continued)

Outcome measure ³	Study sample	Sample size (students/schools)	Author's findings from the study		WWC calculations			
			Mean outcome (standard deviation ²)		Mean difference ⁴ (<i>Caring School Community</i> – comparison)	Effect size ⁵	Statistical significance ⁶ (at $\alpha = 0.05$)	Improvement index ⁷
			<i>Caring School Community</i> group	Comparison group				
The Six-District Study (quasi-experimental design) (continued)								
Stolen money or property	Grades 5–6	635/10	0.26 (0.43)	0.25 (0.45)	–0.01	–0.02	ns	–1
Carried a knife, gun, or other weapon	Grades 5–6	635/10	0.24 (0.49)	0.23 (0.45)	–0.01	–0.02	ns	–1
Threatened to hurt someone	Grades 5–6	635/10	0.45 (0.56)	0.43 (0.57)	–0.02	–0.04	ns	–1
Hurt someone on purpose	Grades 5–6	635/10	0.40 (0.54)	0.37 (0.55)	–0.03	–0.06	ns	–2
Taken a car without permission	Grades 5–6	635/10	0.03 (0.21)	0.09 (0.34)	0.06	0.22	ns	+9
Been in a gang fight	Grades 5–6	635/10	0.10 (0.35)	0.14 (0.40)	0.04	0.11	ns	+4
Thrown objects at people	Grades 5–6	635/10	0.22 (0.44)	0.23 (0.47)	0.01	0.02	ns	+1
Been made fun of or called names	Grades 5–6	635/10	0.98 (0.54)	0.99 (0.57)	0.01	0.02	ns	+1
Had property damaged on purpose	Grades 5–6	635/10	0.44 (0.52)	0.56 (0.56)	0.12	0.22	ns	+9
Had property stolen from desk	Grades 5–6	635/10	0.49 (0.54)	0.67 (0.61)	0.18	0.31	ns	+12
Had money or property taken by force or threat	Grades 5–6	635/10	0.17 (0.40)	0.22 (0.49)	0.05	0.11	ns	+4
Been threatened with harm	Grades 5–6	635/10	0.45 (0.55)	0.47 (0.57)	0.02	0.04	ns	+1
Been physically attacked	Grades 5–6	635/10	0.28 (0.47)	0.34 (0.53)	0.06	0.12	ns	+5
Average ⁸ for behavior (The Six-District Study)						0.07	ns	+3

(continued)

Appendix A3.1 Summary of study findings included in the rating for the behavior domain¹ (continued)

Outcome measure ³	Study sample	Sample size (students/schools)	Author's findings from the study		WWC calculations			
			Mean outcome (standard deviation ²)		Mean difference ⁴ (Caring School Community – comparison)	Effect size ⁵	Statistical significance ⁶ (at $\alpha = 0.05$)	Improvement index ⁷
			Caring School Community group	Comparison group				
Domain average ⁸ for behavior across all studies					0.21	ns	+8	

ns = not statistically significant

1. This appendix reports findings considered for the effectiveness rating and the improvement index. For the San Ramon Study, outcomes in the behavior domain were reported by the study authors for the first cohort only. For the purposes of this review, where findings were reported for multiple points in time for the same sample, only the most recent posttests with an eligible design were reviewed. For the Six-District Study, the findings pertain to outcomes after three years of program implementation.
2. The standard deviation across all students in each group shows how dispersed the participants' outcomes are: a smaller standard deviation on a given measure would indicate that participants had more similar outcomes. The standard deviations for the four behavior outcomes reported by Battistich et al. (1989) were estimated given that the authors used T-scores standardized within each grade level. A T-score is based on a normal distribution with a mean of 50 and standard deviation of 10. The findings included in this domain were eligible for review based on matching of the schools at baseline.
3. Follow-up findings for the study reported by Battistich et al. (1989) were reported by Battistich (2003) for fifth-grade students using peer nominations of negative and positive behaviors. These outcomes were not included in this review because of attrition of schools in the follow-up years.
4. Positive differences and effect sizes favor the intervention group; negative differences and effect sizes favor the comparison group. The direction of the mean difference was reversed for all outcomes except for altruistic behavior in the Battistich et al. (2000) study, so that a positive difference is associated with a decreased frequency of problem behavior.
5. For an explanation of the effect size calculation, please see the [Technical Details of WWC-Conducted Computations](#).
6. Statistical significance is the probability that the difference between groups is a result of chance rather than a real difference between the groups. The level of statistical significance was reported by the study authors or, where necessary, calculated by the WWC to correct for clustering within classrooms or schools and for multiple comparisons. For an explanation about the clustering correction, see the [WWC Tutorial on Mismatch](#). See the [Technical Details of WWC-Conducted Computations](#) for the formulas the WWC used to calculate statistical significance. In the case of both Battistich et al. (1989) and Battistich et al. (2000), corrections for clustering and multiple comparisons were needed, so the statistical significance reported by the WWC may differ from that reported by the authors.
7. The improvement index represents the difference between the percentile rank of the average student in the intervention condition and that of the average student in the comparison condition. The improvement index can take on values between –50 and +50, with positive numbers denoting favorable results.
8. The WWC-computed average effect sizes for each study and for the domain across studies are simple averages rounded to two decimal places. The average improvement indices are calculated from the average effect sizes.

Appendix A3.2 Summary of study findings included in the rating for the knowledge, attitudes, and values domain¹

			Author's findings from the study ²					
			Mean outcome (standard deviation ³)		WWC calculations			
Outcome measure ⁴	Study sample ⁵	Sample size (students/ schools)	<i>Caring School Community group</i>	Comparison group	Mean difference ⁶ (<i>Caring School Community</i> – comparison)	Effect size ⁷	Statistical significance ⁸ (at $\alpha = 0.05$)	Improvement index ⁹
The San Ramon Study (randomized controlled trial with confounding problems)								
Democratic values	Grade 4	294/6	3.31 (0.37)	3.12 (0.45)	0.19	0.46	ns	+18
Perceptual benevolence	Grade 3 (cohort 1)	256/6	1.79 (0.19)	1.75 (0.24)	0.04	0.17	ns	+7
Empathy	Grade 4 (cohort 1)	256/6	1.69 (0.47)	1.67 (0.56)	0.02	na ¹²	ns	na ¹²
Concern for others	Grade 4 (cohort 1)	294/6	2.01 (1.30)	2.06 (1.23)	−0.05	na ¹²	ns	na ¹²
Concern for equality	Grade 4 (cohort 1)	294/6	3.19 (1.19)	2.92 (1.36)	0.27	na ¹²	ns	na ¹²
Motive to help others learn	Grade 4 (cohort 1)	294/6	2.62 (0.92)	2.64 (0.92)	−0.02	na ¹²	ns	na ¹²
Social understanding	Grade 4 (cohort 1)	317/6	2.99 (0.42)	2.93 (0.38)	0.06	0.15	ns	+6
Social problem solving interview	Grade 1 (cohort 2)	295/6	2.21 (na)	2.11 (na)	0.10	0.14	ns	+6
Conflict resolution interview	Kindergarten (cohort 2)	318/6	1.24 (na)	1.08 (na)	0.16	0.18	ns	+7
Social problem solving interview	Grades 1 and 3 (cohort 1)	191/6	2.58 (na)	2.44 (na)	0.15	0.27	ns	+11
Conflict resolution interview	Grades K, 2, and 4 (cohort 1)	133/6	1.36 (na)	1.16 (na)	0.20	0.43	ns	+17
Total self-esteem	Grade 4 (cohort 1)	294/6	2.35 (0.80)	2.40 (0.89)	−0.05	−0.06	ns	−2
Sense of community	Grade 4	294/6	1.54 (0.34)	1.53 (0.38)	0.15	0.39	ns	+15
Average ¹⁰ for knowledge, attitudes, and values (The San Ramon Study)						0.24	ns	+9

(continued)

Appendix A3.2 Summary of study findings included in the rating for the knowledge, attitudes, and values domain¹ (continued)

			Author's findings from the study ²					
			Mean outcome (standard deviation ³)		WWC calculations			
Outcome measure ⁴	Study sample ⁵	Sample size (students/ schools)	<i>Caring School Community group</i>	Comparison group	Mean difference ⁶ (<i>Caring School Community</i> – comparison)	Effect size ⁷	Statistical significance ⁸ (at $\alpha = 0.05$)	Improvement index ⁹
The Six-District Study (quasi-experimental design)								
Democratic values	Grades 3–5	1,265/10	0.13 (0.57)	0.03 (0.57)	0.10	0.18	ns	+7
Acceptance of outgroups	Grades 3–5	1,265/10	0.04 (0.42)	0.00 (0.47)	0.04	0.09	ns	+4
Outgroups discrepancy score (deviation from friend)	Grades 3–5	1,265/10	0.02 (0.48)	–0.09 (0.52)	0.11	0.22	ns	+9
Concern for others	Grades 4–5	568/10	0.02 (0.85)	–0.08 (0.83)	0.10	0.12	ns	+5
Enjoyment of help- ing others learn	Grades 4–5	568/10	–0.04 (0.85)	–0.08 (0.85)	0.04	0.05	ns	+2
Social competence	Grades 4–5	568/10	–0.06 (0.64)	–0.08 (0.71)	0.02	0.03	ns	+1
General self-esteem	Grades 4–5	568/10	0.01 (0.99)	–0.02 (1.01)	0.03	0.03	ns	+1
Conflict resolution skills	Grades 3–6	1,986/6	0.30 (1.04)	0.07 (1.04)	0.23	0.22	ns	+9
Sense of efficacy	Grades 3–6	1,986/6	0.19 (0.67)	0.13 (0.73)	0.06	0.09	ns	+3
Sense of community ¹¹	Grades 3–6	1,986/6	0.09 (0.62)	–0.20 (0.59)	0.29	0.48	ns	+18
Average ¹⁰ for knowledge, attitudes, and values (The Six-District Study)						0.15	ns	+6
Domain average ¹⁰ for knowledge, attitudes, and values across all studies						0.19	na	+8

ns = not statistically significant

na = not applicable

1. This appendix reports findings considered for the effectiveness rating and the improvement index. For the purposes of this review, where findings were reported for multiple points in time for the same sample, only the most recent posttests with an eligible design were reviewed. Subscale findings for the social problem solving interview measure and the conflict resolution interview measure are presented in Appendix A4.1. When averaging the subscales to create these two measures, the effect sizes for the proportion antisocial and prosocial strategies were weighted because those two outcomes were not independent.
2. The WWC obtained from the study authors means and standard deviations for the following outcomes: social problem solving interview scores, conflict resolution interview scores, democratic values, social understanding, perspective (continued)

Appendix A3.2 Summary of study findings included in the rating for the knowledge, attitudes, and values domain¹ (continued)

taking and social competence for the findings of the study reported by Battistich et al. (1989). Some of the outcome measures were administered in multiple years, and the reported outcomes pertain to students who participated in the study during those multiple years. Student-level standard deviations for the following outcome measures were estimated from classroom-level standard deviations: empathy, motives to help others learn, concern for equality, and concern for others.

3. The standard deviation across all students in each group shows how dispersed the participants' outcomes are: a smaller standard deviation on a given measure would indicate that participants had more similar outcomes.
4. Two student outcomes included in the study reported by Battistich et al. (1989), loneliness at school and social anxiety, were not included in this review because they were not considered directly relevant to character education. In addition, the following student outcomes included in the study reported by Battistich et al. (1989) were not included in the review because of lack of psychometric and descriptive information on the measures or lack of statistical information: competitive orientation; helping choices; helping reasons; response to transgressions; transgressions reasons; intrinsic prosocial motivation; and classmates nominations as friends, prosocial, impulsive, competitive, or loners.
5. The study reported by Battistich et al. (1989) used two cohorts of students: cohort 1 started kindergarten in 1982–83 and cohort 2 started in the 1985–86 school year.
6. Positive differences and effect sizes favor the intervention group; negative differences and effect sizes favor the comparison group.
7. For an explanation of the effect size calculation, please see the [Technical Details of WWC-Conducted Computations](#).
8. Statistical significance is the probability that the difference between groups is a result of chance rather than a real difference between the groups. The level of statistical significance was reported by the study authors or, where necessary, calculated by the WWC to correct for clustering within classrooms or schools and for multiple comparisons. For an explanation about the clustering correction, see the [WWC Tutorial on Mismatch](#). See the [Technical Details of WWC-Conducted Computations](#) for the formulas the WWC used to calculate statistical significance. In the case of both Battistich et al. (1989) and Battistich et al. (2000), corrections for clustering and multiple comparisons were needed, so the statistical significance reported by the WWC may differ from that reported by the authors.
9. The improvement index represents the difference between the percentile rank of the average student in the intervention condition and that of the average student in the comparison condition. The improvement index can take on values between –50 and +50, with positive numbers denoting favorable results.
10. The WWC-computed average effect sizes for each study and for the domain across studies are simple averages rounded to two decimal places. The average improvement indices were calculated from the average effect sizes.
11. This measure includes three subscales: students' sense of autonomy in the classroom, classroom supportiveness, and school supportiveness. While two of the subscales measure only student behavior, the third, students' sense of autonomy in the classrooms, may include elements of teacher behavior in addition to student behavior. The WWC did not obtain statistical information on the classroom supportiveness and school supportiveness subscales separately; therefore the composite score was reviewed.
12. Student-level standard deviations were not available in the San Ramon Study for empathy, concern for others, concern for equality, and motive to help others learn. Classroom-level standard deviations were 0.47, 1.30, 1.19, and 0.92 for the intervention group and 0.56, 1.23, 1.36, and 0.92 for the comparison group. Because the student-level effect size and improvement index could not be computed, the magnitude of the effect size of those four measures was not considered for rating purposes. However, the statistical significance for those measures is comparable to other studies and is included in the intervention rating. For further details, please see [Technical Details of WWC-Conducted Computations](#).

Appendix A3.3 Summary of study findings included in the rating for the academic achievement domain¹

Author's findings from the study ²								
			Mean outcome (standard deviation ³)		WWC calculations			
Outcome measure	Study sample ⁴	Sample size (students/ schools)	<i>Caring School Community group</i>	Comparison group	Mean difference ⁵ (<i>Caring School Community</i> – comparison)	Effect size ⁶	Statistical significance ⁷ (at $\alpha = 0.05$)	Improvement index ⁸
The San Ramon Study (randomized controlled trial with confounding problems)								
California Achievement Test—total	Grade 4	339/6	712.16 (na)	712.36 (na)	−0.20	−0.02	ns	−1
Holistic measure of reading comprehension	Grade 6	236/6	51.43 (9.82)	48.02 (9.96)	3.41	0.34	ns	+13
Average ⁹ for academic achievement (The San Ramon Study)						0.16	ns	+6
The Six-District Study (quasi-experimental design)								
Inductive reasoning ¹⁰	Grades 5–6	643/10	2.00 (nr)	1.51 (nr)	0.49	0.03	ns	+1
SAT9—total	Grades 1–5 (Southeastern district)	2675/4	1.38 (na)	4.69 (na)	−3.32	−0.21	ns	−8
SRA Achievement Series—total	Grades 2–6 (West Coast district)	1044/4	−1.52 (na)	−0.70 (na)	−0.82	−0.04	ns	−2
State-developed test	Grade 3 (Southern district)	351/4	0.22 (na)	−0.03 (na)	0.25	0.42	ns	+16
Average ⁹ for academic achievement (The Six-District Study)						0.05	ns	+2
Domain average ⁹ for academic achievement across all studies						0.11	na	+4

ns = not statistically significant

na = not applicable

nr = not reported

1. This appendix reports on findings considered for the effectiveness rating and the improvement index. For the San Ramon Study, outcomes in the academic achievement domain were reported for the first cohort only. Subscale findings for standardized achievement tests are presented in Appendix A4.2. Additional findings of academic motivation are presented in Appendix A4.3.
2. The WWC obtained from the study author the means and standard deviations for all achievement outcomes. The study reported by Battistich et al. (2000) also examined findings for a state-developed achievement test. Those findings were not reviewed by the WWC because of lack of information about the psychometric properties of the test. Although the study author reported statistically significant positive effects on the math, science, and social science subsets of this test, none of these effects were statistically significant (as calculated by the WWC) after correcting for clustering effects at the school level.
3. The standard deviation across all students in each group shows how dispersed the participants' outcomes are: a smaller standard deviation on a given measure would indicate that participants had more similar outcomes.
4. The study reported by Battistich et al. (1989) used two cohorts of students: cohort 1 started kindergarten in 1982–83 and cohort 2 started in the 1985–86 school year.
5. Positive differences and effect sizes favor the intervention group; negative differences and effect sizes favor the comparison group.

Appendix A3.3 Summary of study findings included in the rating for the academic achievement domain¹ *(continued)*

6. For an explanation of the effect size calculation, please see the [Technical Details of WWC-Conducted Computations](#).
7. Statistical significance is the probability that the difference between groups is a result of chance rather than a real difference between the groups. The level of statistical significance was reported by the study authors or, where necessary, calculated by the WWC to correct for clustering within classrooms or schools and for multiple comparisons. For an explanation about the clustering correction, see the [WWC Tutorial on Mismatch](#). See the [Technical Details of WWC-Conducted Computations](#) for the formulas the WWC used to calculate statistical significance. In the case of both the San Ramon Study and the Six-District Study, corrections for clustering and multiple comparisons were needed, so the statistical significance reported by the WWC may differ from that reported by the authors.
8. The improvement index represents the difference between the percentile rank of the average student in the intervention condition and that of the average student in the comparison condition. The improvement index can take on values between -50 and +50, with positive numbers denoting favorable results.
9. The WWC-computed average effect sizes for each study and for the domain across studies are simple averages rounded to two decimal places. The average improvement indices were calculated from the average effect sizes.
10. The effect size for this outcome measure was calculated using the pooled standard deviation. The WWC requested and received from the study authors the pooled standard deviation, which was 15.92 for inductive reasoning.

Appendix A4.1 Summary of findings for knowledge, attitudes, and values¹

			Author's findings from the study					
			Mean outcome (standard deviation ²)		WWC calculations			
Outcome measure	Study sample	Sample size (students/ schools)	Caring School Community group	Comparison group	Mean difference ³ (Caring School Community – comparison)	Effect size ⁴	Statistical significance ⁵ (at $\alpha = 0.05$)	Improvement index ⁶
The San Ramon Study (randomized controlled trial with confounding problems)								
Social problem-solving interview (cohort 1)								
Social problem-solving interview: interpersonal sensitivity	Grades 1 and 3	191/6	2.7 (0.64)	2.48 (0.50)	0.22	0.38	ns	+15
Social problem-solving interview: outcome expectancies	Grades 1 and 3	191/6	4.95 (0.58)	4.81 (0.69)	0.14	0.22	ns	+9
Social problem-solving interview: means-ends cognitive problem solving	Grades 1 and 3	191/6	0.59 (0.40)	0.37 (0.38)	0.22	0.56	ns	+21
Social problem-solving interview: problem resolution strategies	Grades 1 and 3	191/6	4.81 (0.73)	4.73 (0.73)	0.08	0.11	ns	+4
Social problem-solving interview: obstacle problem resolution strategies	Grades 1 and 3	191/6	4.68 (0.77)	4.48 (0.87)	0.20	0.24	ns	+10
Social problem-solving interview: obstacle outcome expectancies	Grades 1 and 3	191/6	4.67 (0.58)	4.34 (0.69)	0.33	0.52	ns	+20
Social problem-solving interview: obstacle means-end cognitive problem solving	Grades 1 and 3	191/6	0.49 (0.40)	0.39 (0.38)	0.10	0.26	ns	+10
Social problem-solving interview: proportion prosocial strategies	Grades 1 and 3	191/6	0.34 (0.13)	0.32 (0.12)	0.02	0.16	ns	+6
Social problem-solving interview: proportion antisocial strategies	Grades 1 and 3	191/6	0.03 (0.05)	0.03 (0.04)	0.00	0.00	ns	+0
Conflict resolution interview (cohort 1)								
Conflict resolution interview: general conflict resolution strategy	Grades K, 2, and 4	133/6	2.76 (0.91)	2.45 (0.86)	0.31	0.35	ns	+14
Conflict resolution interview: proportion prosocial strategies	Grades K, 2, and 4	133/6	0.52 (0.14)	0.44 (0.17)	0.08	0.51	ns	+19

(continued)

Appendix A4.1 Summary of findings for knowledge, attitudes, and values¹ (continued)

			Author's findings from the study					
			Mean outcome (standard deviation ²)		WWC calculations			
Outcome measure	Study sample	Sample size (students/ schools)	<i>Caring School Community group</i>	Comparison group	Mean difference ³ (<i>Caring School Community – comparison</i>)	Effect size ⁴	Statistical significance ⁵ (at $\alpha = 0.05$)	Improvement index ⁶
The San Ramon Study (randomized controlled trial with confounding problems) <i>(continued)</i>								
Conflict resolution interview (cohort 1) <i>(continued)</i>								
Conflict resolution interview: proportion antisocial strategies	Grades K, 2, and 4	133/6	0.10 (0.12)	0.13 (0.13)	−0.03	0.24	ns	+9
Conflict resolution interview: consideration of others' needs	Grades K, 2, and 4	133/6	2.05 (0.74)	1.63 (0.63)	0.42	0.61	ns	+23
Conflict resolution interview (cohort 2)								
Conflict resolution interview: general conflict resolution strategy	Kindergarten	318/6	2.63 (0.87)	2.36 (0.81)	0.27	0.32	ns	+13
Conflict resolution interview: consideration of others' needs	Kindergarten	318/6	1.80 (1.14)	1.56 (0.82)	0.24	0.24	ns	+9
Conflict resolution interview: proportion prosocial strategies	Kindergarten	318/6	0.47 (0.27)	0.33 (0.25)	0.14	0.53	ns	+20
Conflict resolution interview: proportion antisocial strategies	Kindergarten	318/6	0.05 (0.13)	0.08 (0.17)	−0.03	−0.20	ns	−8
Social problem-solving interview (cohort 2)								
Social problem-solving interview: proportion prosocial strategies	Grade 1	295/6	0.38 (0.15)	0.39 (0.16)	−0.01	−0.06	ns	−3
Social problem-solving interview: proportion antisocial strategies	Grade 1	295/6	0.04 (0.08)	0.04 (0.10)	0.00	0.00	ns	+0
Social problem-solving interview: means-ends cognitive problem solving	Grade 1	295/6	0.30 (0.37)	0.20 (0.34)	0.10	0.28	ns	+11
Social problem-solving interview: obstacle problem resolution strategies	Grade 1	295/6	3.98 (1.05)	4.12 (1.11)	−0.14	−0.13	ns	−5
Social problem-solving interview: obstacle outcome expectancies	Grade 1	295/6	4.61 (0.94)	4.45 (1.00)	0.16	0.16	ns	+7

(continued)

Appendix A4.1 Summary of findings for knowledge, attitudes, and values¹ (continued)

Outcome measure	Study sample	Sample size (students/schools)	Author's findings from the study		WWC calculations			
			Mean outcome (standard deviation ²)		Mean difference ³ (<i>Caring School Community</i> – comparison)	Effect size ⁴	Statistical significance ⁵ (at $\alpha = 0.05$)	Improvement index ⁶
			<i>Caring School Community</i> group	Comparison group				
The San Ramon Study (randomized controlled trial with confounding problems) <i>(continued)</i>								
Social problem-solving interview (cohort 2) <i>(continued)</i>								
Social problem-solving interview: obstacle means-end cognitive problem solving	Grade 1	295/6	0.58 (0.63)	0.47 (0.53)	0.11	0.19	ns	+7
Social problem-solving interview: interpersonal sensitivity	Grade 1	295/6	2.09 (0.40)	2.03 (0.36)	0.06	0.16	ns	+6
Social problem-solving interview: primary resolution strategies	Grade 1	295/6	3.67 (0.89)	3.45 (0.82)	0.22	0.26	ns	+10

ns = not statistically significant

1. This appendix presents findings for measures of academic motivation. Because these outcomes do not represent actual academic performance or persistence, they are not considered for rating purposes. However, because those outcomes may be linked to academic performance or persistence they are presented here.
2. The standard deviation across all students in each group shows how dispersed the participants' outcomes are: a smaller standard deviation on a given measure would indicate that participants had more similar outcomes.
3. Positive differences and effect sizes favor the intervention group; negative differences and effect sizes favor the comparison group.
4. For an explanation of the effect size calculation, please see the [Technical Details of WWC-Conducted Computations](#).
5. Statistical significance is the probability that the difference between groups is a result of chance rather than a real difference between the groups. The level of statistical significance was reported by the study authors or, where necessary, calculated by the WWC to correct for clustering within classrooms or schools (corrections for multiple comparisons were not done for findings not included in the overall intervention rating). For an explanation about the clustering correction, see the [WWC Tutorial on Mismatch](#). See the [Technical Details of WWC-Conducted Computations](#) for the formulas the WWC used to calculate statistical significance. In the case of Battistich et al. (1989) and Battistich et al. (2000), corrections for clustering were needed, so the statistical significance reported by the WWC may differ from that reported by the authors.
6. The improvement index represents the difference between the percentile rank of the average student in the intervention condition and that of the average student in the comparison condition. The improvement index can take on values between –50 and +50, with positive numbers denoting favorable results.

Appendix A4.2 Summary of subtest study findings for the academic achievement domain¹

Author's findings from the study ²								
Outcome measure	Study sample ⁴	Sample size (students/schools)	Mean outcome (standard deviation ³)		WWC calculations			
			Caring School Community group	Comparison group	Mean difference ⁵ (Caring School Community – comparison)	Effect size ⁶	Statistical significance ⁷ (at $\alpha = 0.05$)	Improvement index ⁸
The San Ramon Study (randomized controlled trial with confounding problems)								
California Achievement Test—Reading	Grade 4	348/6	730.07 (33.53)	733.33 (60.96)	−3.26	−0.07	ns	−3
California Achievement Test—Language	Grade 4	349/6	713.6 (29.67)	715.85 (30.15)	−2.25	−0.08	ns	−3
California Achievement Test—Math	Grade 4	346/6	728.04 (29.32)	730.93 (25.38)	−2.89	−0.10	ns	−4
California Achievement Test—Word analysis	Grade 4	287/6	735.53 (76.37)	729.95 (75.27)	5.58	0.07	ns	+3
California Achievement Test—Spelling	Grade 4	347/6	713.5 (30.87)	710.66 (30.59)	2.84	0.09	ns	+4
California Achievement Test—Study skills	Grade 4	344/6	728.37 (30.44)	728.30 (33.52)	0.07	0.00	ns	+0
California Achievement Test—Science	Grade 4	345/6	667.64 (4.23)	666.91 (43.09)	0.73	0.02	ns	+1
California Achievement Test—Social studies	Grade 4	345/6	680.54 (25.81)	682.96 (22.48)	−2.42	−0.10	ns	−4
Holistic measure of reading comprehension	Grade 6	236/6	51.43 (9.82)	48.02 (9.96)	3.41	0.34	ns	+13
The Six-District Study (quasi-experimental design)								
SAT9—Reading ⁹	Grades 1–5 (Southeastern district)	2,675/4	1.32 (nr)	3.13 (nr)	−1.81	−0.13	ns	−5
SAT9—Math ⁹	Grades 1–5 (Southeastern district)	2,675/4	1.43 (nr)	6.25 (nr)	−4.82	−0.29	ns	−11

(continued)

Appendix A4.2 Summary of subtest study findings for the academic achievement domain¹ (continued)

Outcome measure	Study sample ⁴	Sample size (students/schools)	Author's findings from the study ²		WWC calculations			
			Mean outcome (standard deviation ³)		Mean difference ⁵ (<i>Caring School Community</i> – comparison)	Effect size ⁶	Statistical significance ⁷ (at $\alpha = 0.05$)	Improvement index ⁸
			<i>Caring School Community</i> group	Comparison group				
The Six-District Study (quasi-experimental design) <i>(continued)</i>								
SRA Achievement Series—Reading ⁹	Grades 2–6 (West Coast district)	1,044/4	2.00 (nr)	2.00 (nr)	0.00	0.00	ns	+0
SRA Achievement Series—Math ⁹	Grades 2–6 (West Coast district)	1,044/4	–5.03 (nr)	–3.39 (nr)	–1.64	–0.08	ns	–3
Reading (state-developed test)	Grade 3 (Southern district)	351/4	0.06 (nr)	–0.01 (nr)	0.07	0.13	ns	+5
Math (state-developed test)	Grade 3 (Southern district)	351/4	0.43 (nr)	–0.13 (nr)	0.56	0.90	ns	+32
Science (state-developed test)	Grade 3 (Southern district)	351/4	0.25 (nr)	0.04 (nr)	0.21	0.40	ns	+15
Social science (state-developed test)	Grade 3 (Southern district)	351/4	0.14 (nr)	–0.01 (nr)	0.15	0.24	ns	+9

ns = not statistically significant

nr = not reported

1. This appendix reports on subtest findings not considered for the effectiveness rating and the improvement index. For the San Ramon Study, outcomes in the academic achievement domain were reported for the first cohort only.
2. The WWC obtained from the study author the means and standard deviations for all achievement outcomes. The study reported by Battistich et al. (2000) also examined findings for a state-developed achievement test. Those findings were not reviewed by the WWC because of lack of information about the psychometric properties of the test. Although the study author reported statistically significant positive effects on the math, science, and social science subsets of this test, none of these effects were statistically significant (as calculated by the WWC) after correcting for clustering effects at the school level.
3. The standard deviation across all students in each group shows how dispersed the participants' outcomes are: a smaller standard deviation on a given measure would indicate that participants had more similar outcomes.
4. The study reported by Battistich et al. (1989) used two cohorts of students: cohort 1 started kindergarten in 1982–83 and cohort 2 started in the 1985–86 school year.
5. Positive differences and effect sizes favor the intervention group; negative differences and effect sizes favor the comparison group.
6. For an explanation of the effect size calculation, please see the [Technical Details of WWC-Conducted Computations](#).
7. Statistical significance is the probability that the difference between groups is a result of chance rather than a real difference between the groups. The level of statistical significance was reported by the study authors or, where necessary, calculated by the WWC to correct for clustering within classrooms or schools (corrections for multiple comparisons were not done for findings not included in the overall intervention rating). For an explanation about the clustering correction, see the [WWC Tutorial on Mismatch](#). See the [Technical Details of WWC-Conducted Computations](#) for the formulas the WWC used to calculate statistical significance. In the case of both Battistich et al. (1989) and Battistich et al. (2000), corrections for clustering were needed, so the statistical significance reported by the WWC may differ from that reported by the authors.
8. The improvement index represents the difference between the percentile rank of the average student in the intervention condition and that of the average student in the comparison condition. The improvement index can take on values between –50 and +50, with positive numbers denoting favorable results.
9. The effect size for this outcome measure was calculated using the pooled standard deviation. The WWC requested and received from the study authors the pooled standard deviations, which were 14.05 for SAT9—Reading, 16.51 for SAT9—Math, 18.6 for SRA—Reading, and 20.23 for SRA—Math. Because districts included in this study varied in the standardized tests used (all academic measures except for the inductive reasoning test), these findings pertain to two CSC schools and their matched comparison schools.

Appendix A4.3 Summary of additional findings for academic achievement¹

Outcome measure	Study sample	Sample size (students/ schools)	Author's findings from the study		WWC calculations			
			Mean outcome (standard deviation ²)		Mean difference ³ (<i>Caring School Community</i> – comparison)	Effect size ⁴	Statistical significance ⁵ (at $\alpha = 0.05$)	Improvement index ⁶
			<i>Caring School Community</i> group	Comparison group				
The San Ramon Study (randomized controlled trial with confounding problems)								
Achievement motivation	Grade 4 (cohort 1)	294/6	1.65 (0.17)	1.68 (0.19)	0.01	0.05	ns	+2
The Six-District Study (quasi-experimental design)								
Task orientation	Grades 3–6	1,986/6	0.05 (0.85)	–0.16 (0.88)	0.21	0.24	ns	+10
Ego orientation	Grades 3–6	1,986/6	0.20 (1.12)	0.17 (1.12)	0.03	0.03	ns	+1
Work avoidance	Grades 3–6	1,986/6	0.12 (1.06)	0.18 (1.06)	–0.06	–0.06	ns	–2
Intrinsic academic motivation	Grades 3–4	1,986/10	2.63 (5.65)	0.76 (5.65)	1.87	0.33	ns	+13
Preference for challenging tasks	Grades 3–4	660/10	0.02 (0.28)	0.03 (0.08)	–0.01	–0.05	ns	–2
Frequency reading self-chosen books outside of school	Grades 3–6	1,986/10	–0.12 (1.36)	–0.27 (1.38)	0.15	0.11	ns	+4
Frequency reading self-chosen books in school	Grades 3–6	1,986/10	0.07 (0.47)	–0.11 (0.53)	0.18	0.36	ns	+14
Educational aspirations	Grades 3–6	1,986/10	0.08 (0.74)	0.05 (0.84)	0.03	0.04	ns	+2
Educational expectations	Grades 3–6	1,986/10	0.13 (0.76)	0.18 (0.82)	–0.05	–0.06	ns	–3
Academic self-esteem	Grades 3–6	1,986/10	0.22 (0.84)	0.13 (0.80)	0.09	0.11	ns	+4
Liking for reading	Grades 3–6	1,986/10	–0.09 (0.98)	–0.16 (0.95)	0.07	0.07	ns	+3

ns = not statistically significant

(continued)

Appendix A4.3 Summary of additional findings for academic achievement¹ *(continued)*

1. This appendix presents findings for measures that fall in the academic achievement domain but are measures of academic motivation rather than direct measures of achievement.
2. The standard deviation across all students in each group shows how dispersed the participants' outcomes are: a smaller standard deviation on a given measure would indicate that participants had more similar outcomes.
3. Positive differences and effect sizes favor the intervention group; negative differences and effect sizes favor the comparison group.
4. For an explanation of the effect size calculation, please see the [Technical Details of WWC-Conducted Computations](#).
5. Statistical significance is the probability that the difference between groups is a result of chance rather than a real difference between the groups. The level of statistical significance was reported by the study authors or, where necessary, calculated by the WWC to correct for clustering within classrooms or schools (corrections for multiple comparisons were not done for findings not included in the overall intervention rating). For an explanation about the clustering correction, see the [WWC Tutorial on Mismatch](#). See the [Technical Details of WWC-Conducted Computations](#) for the formulas the WWC used to calculate statistical significance. In the case of Battistich et al. (1989) and Battistich et al. (2000), corrections for clustering were needed, so the statistical significance reported by the WWC may differ from that reported by the authors.
6. The improvement index represents the difference between the percentile rank of the average student in the intervention condition and that of the average student in the comparison condition. The improvement index can take on values between -50 and +50, with positive numbers denoting favorable results.

Appendix A5.1 *Caring School Community™* rating for the behavior domain

The WWC rates the effects of an intervention in a given outcome domain as: positive, potentially positive, mixed, no discernible effects, potentially negative, or negative.¹

For the outcome domain of behavior, the WWC rated *Caring School Community™* as having potentially positive effects. It did not meet the criteria for positive effects because no studies met WWC evidence standards for a strong design. The remaining ratings (mixed effects, no discernible effects, potentially negative effects, and negative effects) were not considered because *Caring School Community™* was assigned the highest applicable rating.

Rating received

Potentially positive effects: Evidence of a positive effect with no overriding contrary evidence.

- Criterion 1: At least one study showing a statistically significant or substantively important *positive* effect.

Met. One study showed a statistically significant positive effect. In addition, the average effect size for this study was large enough to be considered substantively important, according to WWC criteria.

- Criterion 2: No studies showing a statistically significant or substantively important *negative* effect and fewer or the same number of studies showing *indeterminate* effects than showing statistically significant or substantively important *positive* effects.

Met. No studies showed a statistically significant or substantively important negative effect. One study showed indeterminate effects and one study showed statistically significant positive effects.

Other ratings considered

Positive effects: Strong evidence of a positive effect with no overriding contrary evidence.

- Criterion 1: Two or more studies showing statistically significant *positive* effects, at least one of which met WWC evidence standards for a strong design.

Not met. No studies met WWC evidence standards for a strong design.

- Criterion 2: No studies showing statistically significant or substantially important *negative* effects.

Met. No studies showed statistically significant or substantively important negative effects.

1. For rating purposes, the WWC considers the statistical significance of individual outcomes and the domain-level effects. The WWC also considers the size of the domain-level effects for ratings of potentially positive or potentially negative effects. See the [WWC Intervention Rating Scheme](#) for a complete description.

Appendix A5.2 *Caring School Community™* rating for the knowledge, attitudes, and values domain

The WWC rates the effects of an intervention in a given outcome domain as: positive, potentially positive, mixed, no discernible effects, potentially negative, or negative.¹

For the outcome domain of knowledge, attitudes, and values, the WWC rated *Caring School Community™* as having no discernible effects. It did not meet the criteria for other ratings (positive effects, potentially positive effects, mixed effects, potentially negative effects, and negative effects) because the two studies that met WWC standards with reservations did not show statistically significant or substantively important effects.

Rating received

No discernible effects: No affirmative evidence of effects.

- Criterion 1: None of the studies shows a statistically significant or substantively important effect, either *positive* or *negative*.

Met. The two studies that assessed outcomes in this domain both showed indeterminate effects.

Other ratings considered

Positive effects: Strong evidence of a positive effect with no overriding contrary evidence.

- Criterion 1: Two or more studies showing statistically significant *positive* effects, at least one of which met WWC evidence standards for a strong design.

Not met. No studies met WWC evidence standards for a strong design.

- Criterion 2: No studies showing statistically significant or substantively important *negative* effects.

Met. Both studies on *Caring School Community™* showed indeterminate effects.

Potentially positive effects: Evidence of a positive effect with no overriding contrary evidence.

- Criterion 1: At least one study showing a statistically significant or substantively important *positive* effect.

Not met. No studies showed a statistically significant or substantively important positive effect.

- Criterion 2: No studies showing a statistically significant or substantively important *negative* effect and fewer or the same number of studies showing *indeterminate* effects than showing statistically significant or substantively important *positive* effects.

Not met. Both studies on *Caring School Community™* showed indeterminate effects.

Mixed effects: Evidence of inconsistent effects as demonstrated through either of the following criteria.

- Criterion 1: At least one study showing a statistically significant or substantively important *positive* effect, and at least one study showing a statistically significant or substantively important *negative* effect, but no more such studies than the number showing a statistically significant or substantively important *positive* effect.

Not met. No studies showed a statistically significant or substantively important effect, either positive or negative.

- Criterion 2: At least one study showing a statistically significant or substantively important effect, and more studies showing an *indeterminate* effect than showing a statistically significant or substantively important effect.

Not met. No studies showed a statistically significant or substantively important effect.

(continued)

Appendix A5.2 Caring School Community™ rating for the knowledge, attitudes, and values domain (continued)

Potentially negative effects: Evidence of a negative effect with no overriding contrary evidence

- Criterion 1: At least one study showing a statistically significant or substantively important *negative* effect.

Not met. No studies showed a statistically significant or substantively important negative effect.

- Criterion 2: No studies showing a statistically significant or substantively important *positive* effect, or more studies showing statistically significant or substantively important *negative* effects than showing statistically significant or substantively important *positive* effects.

Met. No studies showed a statistically significant or substantively important positive effect.

Negative effects: Strong evidence of a negative effect with no overriding contrary evidence.

- Criterion 1: Two or more studies showing statistically significant *negative* effects, at least one of which met WWC evidence standards for a strong design.

Not met. No studies showed a statistically significant negative effect, and no studies met WWC evidence standards for a strong design.

- Criterion 2: No studies showing statistically significant or substantively important *positive* effects.

Met. No studies showed a statistically significant or substantively important positive effect.

1. For rating purposes, the WWC considers the statistical significance of individual outcomes and the domain-level effects. The WWC also considers the size of the domain-level effects for ratings of potentially positive and potentially negative effects. See the [WWC Intervention Rating Scheme](#) for a complete description.

Appendix A5.3 *Caring School Community™* rating for the academic achievement domain

The WWC rates the effects of an intervention in a given outcome domain as: positive, potentially positive, mixed, no discernible effects, potentially negative, or negative.¹

For the outcome domain of academic achievement, the WWC rated *Caring School Community™* as having no discernible effects. It did not meet the criteria for other ratings (positive effects, potentially positive effects, mixed effects, potentially negative effects, and negative effects) because the two studies that met WWC standards with reservations did not show statistically significant or substantively important effects.

Rating received

No discernible effects: No affirmative evidence of effects.

- Criterion 1: None of the studies shows a statistically significant or substantively important effect, either *positive* or *negative*.

Met. Both studies that assessed outcomes in this domain showed indeterminate effects.

Other ratings considered

Positive effects: Strong evidence of a positive effect with no overriding contrary evidence.

- Criterion 1: Two or more studies showing statistically significant *positive* effects, at least one of which met WWC evidence standards for a strong design.

Not met. No studies met WWC evidence standards for a strong design.

- Criterion 2: No studies showing statistically significant or substantively important *negative* effects.

Met. Both studies on *Caring School Community™* showed indeterminate effects.

Potentially positive effects: Evidence of a positive effect with no overriding contrary evidence.

- Criterion 1: At least one study showing a statistically significant or substantively important *positive* effect.

Not met. No studies showed a statistically significant or substantively important positive effect.

- Criterion 2: No studies showing a statistically significant or substantively important *negative* effect and fewer or the same number of studies showing *indeterminate* effects than showing statistically significant or substantively important *positive* effects.

Not met. Both studies on *Caring School Community™* showed indeterminate effects.

Mixed effects: Evidence of inconsistent effects as demonstrated through either of the following criteria.

- Criterion 1: At least one study showing a statistically significant or substantively important *positive* effect, and at least one study showing a statistically significant or substantively important *negative* effect, but no more such studies than the number showing a statistically significant or substantively important *positive* effect.

Not met. No studies showed a statistically significant or substantively important effect, either positive or negative.

- Criterion 2: At least one study showing a statistically significant or substantively important effect, and more studies showing an *indeterminate* effect than showing a statistically significant or substantively important effect.

Not met. No studies showed a statistically significant or substantively important effect.

(continued)

Appendix A5.3 *Caring School Community™* rating for the academic achievement domain *(continued)*

Potentially negative effects: Evidence of a negative effect with no overriding contrary evidence

- Criterion 1: At least one study showing a statistically significant or substantively important *negative* effect.

Not met. No studies showed a statistically significant or substantively important negative effect.

- Criterion 2: No studies showing a statistically significant or substantively important *positive* effect, or more studies showing statistically significant or substantively important *negative* effects than showing statistically significant or substantively important *positive* effects.

Met. No studies showed a statistically significant or substantively important positive effect.

Negative effects: Strong evidence of a negative effect with no overriding contrary evidence.

- Criterion 1: Two or more studies showing statistically significant *negative* effects, at least one of which met WWC evidence standards for a strong design.

Not met. No studies showed a statistically significant negative effect, and no studies met WWC evidence standards for a strong design.

- Criterion 2: No studies showing statistically significant or substantively important *positive* effects.

Met. No studies showed a statistically significant or substantively important positive effect.

1. For rating purposes, the WWC considers the statistical significance of individual outcomes and the domain-level effects. The WWC also considers the size of the domain-level effects for ratings of potentially positive and potentially negative effects. See the [WWC Intervention Rating Scheme](#) for a complete description.

Appendix A6 Extent of evidence by domain

Outcome domain	Number of studies	Sample size		Extent of evidence ¹
		Schools	Students	
Behavior	2	16	2,336	Moderate to large
Knowledge, attitudes, and values	2	16	2,280	Moderate to large
Academic achievement	2	16	3,719	Moderate to large

na = not applicable/not studied

1. A rating of “moderate to large” requires at least two studies and two schools across studies in one domain, and a total sample size across studies of at least 350 students or 14 classrooms. Otherwise, the rating is “small.”